INVESTIGATING THE EFFECTS OF ENSEMBLE CLASSIFICATION ON REMOTELY SENSED DATA FOR LAND COVER MAPPING

Bolanle Abe[*a], *Anthony Gidudu*[a], *and Tshilidzi Marwala*[b]

[a]School of Electrical & Information Engineering, University of the Witwatersrand, Johannesburg. South Africa.
[b]Faculty of Engineering & the Built Environment, University of Johannesburg. South Africa.
Correspondence: Abe_tolulope@yahoo.com

1. INTRODUCTION

In pattern recognition, land cover information is known to be one important data component for various applications among which are: environmental analysis, global change investigations and in large scale resource management [1], [2]. Remote sensing technology application has been a major tool for deriving land cover information. One major approach of obtaining land cover information from remote sensing data is through classification. The classification accuracy problem related to land cover mapping has been an issue of concern in pattern recognition. The problems include: (i) the proportion between the number of training samples and the available number of bands, and (ii) high correlation between the training patterns from the same area which reduces the information communicated by the sample under consideration to the algorithm [3]. In machine learning, several classification algorithms have been developed for processing remotely sensed data since the first Landsat image was obtained in early 1970s [1], [4]. The purpose is to find ways of improving classification accuracy. Superiority of different approaches employed for classification depends on the selected data set and the strategy used for the designing phase of each classifier. Ensemble classification is a recently developed paradigm in machine learning aimed at improving classification accuracy over single algorithm. In ensemble classification, the constituent classifiers must err in differently so as to introduce diversity into the classification system. Among various methods used to make an ensemble diverse, feature selection approach has yielded considerable good results in terms of classification accuracy [5], [6], [7]. In this study, feature selection techniques are exploited to create diversity in ensemble classification. The methods of combining the predictions of the base classifiers into an integrated ensemble classification result include: majority voting, weighted majority voting, consensus theory and stacking [8], [9], [10]. For this investigation, majority voting was adopted as the consensus rule for combining base classifiers. Using Landsat imagery of Uganda in East Africa and AVIRIS NW Indiana pines, the study focuses on investigating the effects of ensemble size and feature selection on classification accuracy of land cover mapping.

2. REVIEW ON ENSEMBLE

Ensemble technology has been successfully used in many applications including; image classification (e.g. character recognition), handwriting recognition, speech recognition, and fingerprints recognition [8], [9], [10], [11]. In pattern recognition applications of ensemble method has been proved to be an effective tool for increasing classification accuracy. Traditionally, diversity is introduced into an ensemble system through using different classifiers or by training a classifier with different training datasets, such as boosting and bagging [12]. The feature (band) selection approach aims at selecting a subset of the available features that: (1) promote effective diversity between the classes under consideration and (2) have bands that are as most invariant as possible in the spatial field [3], [7].

Feature selection methods help in reducing the processing period required by the classification process and also improve classification accuracy. By varying the feature subsets used to create the ensemble classifier, diversity is ascertained since the base classifiers tend to err in different subspaces of the instance space [13].

Various methods used for selecting features for ensemble systems include; exhaustive search, random selection of feature subsets [14], and genetic algorithms [7]. For the purpose of this study, we explored feature selection approach on remotely sensed imagery using Bhattarcharrya distance, Transformed divergence and Divergence separability index to create diversity in the ensemble method. The generated feature subsets from the datasets were used to train and test the base classifiers.

3. METHODOLOGY

Two different remotely sensed datasets were used for this investigation. The first dataset included a Landsat image of Kampala the capital of Uganda, East Africa. The dataset was generated in 2001 with column 171 and row 60. The image contains five land cover classes which are; water, built up areas, thick swamps, light swamps and other vegetations. Three separability indices were used to select the features for creating the base classifiers. The indices are: Bhattacharyya distance, Divergence and Transformed divergence. For each separability index, five base classifiers were generated using the feature selected for the ensembles. The second dataset included the Indian pines site in North-western Indiana hyperspectral data from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [15]. The dataset designated 92AV3C contains 224 bands out of which 24 bands were discarded because they were affected by atmospheric conditions. 180 bands out of the remaining 200 were used for this investigation. The data set is accompanied with the ground truth information having sixteen labelled land cover classes. The land cover classes are: Alfalfa, building-grass-tree drives, corn-minimum till, corn-notil, corn, grass/pasture, grass/pasture-mowed, grass trees, hay-windrowed, oats, soybeans-minimum till, soybean-notill, soybean-clean, soybean-clean, stone-steel towers, wheat and woods. Bhattacharyya distance separability index was used for exhaustive feature selection search. The research design involved incrementally increasing the number of base classifiers on one hand, and on the other hand, increasing the number of bands per base classifier. Base classifiers were increased from three (3) to ten (10) per ensemble. The number of bands per base classifier was incrementally increased in the following sequence: 2,4,6,8, 10, 12 and 14.

Support Vector machines (SVM) which has been found competitive with the best existing machine learning algorithms in classifying remotely sensed datasets was used to create the ensembles. The SVM is a supervised classification approach which uses optimization algorithms to find the optimal decision boundaries between land cover classes and only small training samples are needed [1], [5]. Researchers employing SVM for pattern recognition have concluded that SVM provides significant advantages in accuracy, simplicity and robustness [16]. Based on the discovered advantages in the application of SVM, Gaussian support vector machine was used for the study.

4. RESULTS, DISCUSSION AND CONCLUSION.

The results obtained from the base classifiers and the ensembles were calculated in terms of Kappa Coefficient of Agreement [17] by comparing the generated maps with the ground truth data. Tables 1and 2 show the results obtained using Kampala and Indiana datasets respectively.

Table 1: Summary of the results obtained using Kampala dataset with the application of 3SI.

SI	Bhattacharyya Distance						Divergence					
N/F	2		3		4		2		3		4	
		BC		BC		BC		BC		BC		BC
BS	5, 6	0.8565	3, 5, 6	0.8718	3, 4, 5, 6	0.9269	2, 4	0.7921	2, 3, 4	0.8964	3, 4, 5, 6	0.9269
	3, 5	0.8496	1, 2, 3	0.3841	2, 3, 5, 6	0.8794	1, 5	0.7283	1, 2, 4	0.8855	2, 4, 5, 6	0.9246
	2, 3	0.5395	4, 5, 6	0.9191	1, 2, 3, 5	0.8599	1, 4	0.7750	4, 5, 6	0.9191	1, 4, 5, 6	0.9211
	4, 5	0.9030	3, 4, 5	0.9201	2, 3, 4, 5	0.9295	3, 4	0.8814	1, 3, 4	0.8574	1, 3, 5, 6	0.8667
	1, 3	0.2220	2, 3, 5	0.8591	1, 2, 3, 4	0.8994	2, 5	0.7569	2, 4, 5	0.9236	2, 3, 5, 6	0.8794
E	**0.8922**		**0.8940**		**0.9160**		**0.8559**		**0.9007**		**0.9237**	
SB	0.9030		0.9201		0.9295		0.8814		0.9236		0.9269	
D	0.42		0.57		0.70		0.47		0.70		0.72	

(a)

SI	Transformed Divergence					
F	2		3		4	
				BC		BC
B	2, 4	0.7921	1, 2, 4	0.8855	1, 2, 5, 6	0.8678
	3, 4	0.8599	2, 3, 4	0.8964	1, 3, 5, 6	0.8667
	4, 6	0.9017	4, 5, 6	0.9191	2, 3, 5, 6	0.8794
	1, 4	0.7750	2, 4, 6	0.9175	3, 4, 5, 6	0.9269
	5, 6	0.8565	1, 3, 4	0.8574	1, 4, 5, 6	0.9211
E	**0.8680**		**0.8982**		**0.9049**	
SB	0.9017		0.9191		0.9269	
D	0.44		0.71		0.65	

(b)

Table 2: Results obtained using Bhattacharyya SI feature selection method Indiana dataset and

	Number of features per Base Classifier												
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
BC1	0.464	0.495	0.563	0.587	0.616	0.614	0.596	0.606	0.621	0.632	0.645	0.619	0.621
BC2	0.470	0.510	0.565	0.579	0.613	0.618	0.613	0.616	0.613	0.621	0.626	0.637	0.632
BC3	0.466	0.522	0.563	0.577	0.566	0.612	0.321	0.615	0.611	0.620	0.619	0.626	0.626
BC4	0.456	0.517	0.563	0.581	0.592	0.618	0.618	0.610	0.623	0.598	0.634	0.611	0.642
BC5	0.484	0.535	0.560	0.581	0.614	0.614	0.616	0.578	0.613	0.624	0.642	0.645	0.632
BC6	0.490	0.528	0.541	0.580	0.603	0.613	0.625	0.620	0.615	0.623	0.616	0.639	0.632
BC7	0.490	0.504	0.558	0.579	0.614	0.626	0.627	0.604	0.623	0.596	0.625	0.647	0.628
BC8	0.489	0.477	0.538	0.568	0.590	0.614	0.615	0.590	0.619	0.593	0.646	0.615	0.624
BC9	0.465	0.494	0.561	0.579	0.593	0.619	0.618	0.615	0.620	0.622	0.628	0.637	0.617
BC10	0.475	0.499	0.560	0.562	0.613	0.614	0.627	0.600	0.634	0.623	0.619	0.623	0.644
E	**0.483**	**0.513**	**0.563**	**0.581**	**0.606**	**0.618**	**0.620**	**0.609**	**0.621**	**0.620**	**0.628**	**0.634**	**0.631**
SB	**0.490**	**0.535**	**0.565**	**0.587**	**0.616**	**0.626**	**0.627**	**0.620**	**0.634**	**0.632**	**0.646**	**0.647**	**0.644**

Note: BC = Base classifier, E = Ensemble, SB = Single Best classifier, SI= Separability Index, N/F = Number of Features per BC, BS = Bands selected, D = Diversity

Comparing the results obtained from the Tables, in all cases, classification accuracy increased as the number of features in the ensemble increase. The results in Table 1 also reflect the measure of diversity per ensemble according to the degree of agreement and variance. It is observed from the results that the ensemble result in each instance is better than majority of the base classifier's predictions, though the best prediction obtained from the base classifier constituting each ensemble is better than the ensemble result. The diversity in the predictions of the base classifiers constituting the ensemble reduces the risk of making a particular poor selection [10], [18].

To further authenticate the results, Binomial Test of Significance (BTS) was conducted on the ensemble results to give a better appreciation of the difference between the ensembles. This is to ascertain the pair wise difference between two ensemble classifications (2 sides test). BTS was carried out at 95% confidence interval such that the null hypothesis (H_o) stating that there is no significant difference will be rejected if $|Z| > 1.96$ [17]. Results obtained reveals that, for ensemble feature classification application for land cover mapping there is no significant benefit in having many base classifiers. From the results, the minimum number of base classifiers is sufficient.

5. REFERENCES

[1] C.Huang, L.S. Davis and J.R.G. Townsend,"An assessment of support vector machines for land cover classification," *Int.Journal of Remote Sensing,* vol. 4, pp 725-749, 2002.

[2] B.M.Steel and D.A. Patterson, 'Land cover mapping using combination and ensemble classifiers', Computing Science and Statistics, vol.33, pp236 – 247, 2001.

[3] L. Bruzzone, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," IEEE *Transactions on Geoscience & Remote Sensing,*vol. 47, no.9, pp.3180 – 3191, sept. 2009.

[4] F.G. Hall, J. R. Townsend and E. T. Engman, "Status of remote sensing algorithms for estimation of land surface state parameters," *Remote Sensing of Environment*, vol, 51, pp. 138 – 156, 1995.

[5] R. Archibald and G. Fann, "Feature selection and classification of Hyperspectral images with support vector machines," *IEEE Geoscience & Remote Sensing Letters*, vol. 4, no. 4, 2007.

[6] R. Huber and L.V. Dutra, "Classifier combination and feature selection for land-cover mapping from high-resolution airborne dual-band SAR data," vol. 5, part 1, pp. 370-375, Orlando, USA, 2000.

[7] D.Optiz, "Feature selection for Ensembles," *In Proceedings of the 16th National Conference on Artificial/Intelligence (AAAI), Orlando-Florida, USA,* pp. 379 – 384, 1999.

[8] S. Tulyakov, S. Jaeger and V. Govindaraju, and D. Doermann, "Review of classifier combination methods," *Studies in Computational Intelligence*, vol. 90, pp. 361 – 386, 2008.

[9] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine,* third quarter, pp. 21 – 45, 2006

[10] J. Kittler, M. Hatef, R.P. Duin, and J. Mates,, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, no.3, pp.226-239, 1998.

[11] L. Xu, A. Krzyzak,C.Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on System, Man, and Cybernetics,* vol. 23, no.3, pp. 418 – 435, 1992.

[12] L. Breiman, "Bagging predictors," *Machine learning,* vol. 26, no. 2, pp.123 – 140,1990.

[13] A. Tsymbal, M. Pechenizkiy and P. Cunnungham, "Diversity in search strategies for ensemble feature selection", *Information Fusion,*vol 6, no.1, pp. 83 – 98, 2005.

[14] T.K. Ho, "The random subspace method for constructing decision forests," *IEEE Trnscation of Pattern Analysis and Machine Intelligence,* vol.20, no.8, pp. 832 – 844, 1998.

[15] AVIRIS NW Indiana's Indian Pines 1992 data set [Online]:
ftp://ftp.ecn.purdue.edu/biehl/Multispec/92AV3C (Original file) and
ftp://ftp.ecn.purdue.edu/biehl/Multispec/ThyFiles.zip (ground truth).

[16] G. Camps-Valls and L. Bruzzone, "Kernel based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 43, no. 6, pp. 1351 – 1362, June 2005.

[17] G.H. Rosenfield and K. Fitzpatrick-Lins, "A coefficient of agreement as a measure of thematic classification accuracy," *Photogrammetric Engineering and Remote Sensing,*vol. 52, pp. 223-227, 1986.